Data Mining CSE 572- Assignment 5 / Mini Project 3 Submission Report

Ankit Nadig (1211213650), anadig@asu.edu
Vraj Delhivala (1211213637), vdelhiva@asu.edu

Problem Statement- Study the application of multi-label learning and ensemble learning on real-world applications

Problem 1

a. We use a SVM classifier with a polynomial kernel and parameter 2 to classify data.into respective classes.
   i. We use the 'fitcsvm' function to train different SVM for every unique class (10 in this case) present in the dataset using kernel function as polynomial and order as 2.
   ii. The SVM Models are structs that are stored in an array.
   iii. Then, the 'predict' function is used on the testData and every SVM Model and a score matrix is computed.
   iv. The SVM for a particular class that gives the highest prediction score, it's class is chosen as the class to associate with the input. Using this we form a label matrix.
   v. The label matrix is then compared to the y_test to find Accuracy
   vi. Accuracy Obtained = 64.3513%

b. We use SVM classifiers to classify and predict data into respective classes.
   i. We use the 'fitcsvm' function to train different SVM for every unique class present in the dataset using kernel function as gaussian and scale as Auto. The SVM Models are structs that are stored in an array.
   ii. Then, the 'predict' function is used on the testData and every SVM Model and a score matrix is computed.
   iii. The SVM for a particular class that gives the highest prediction score, it's class is chosen as the class to associate with the input. Using this we form a label matrix.
   iv. The label matrix is then compared to the y_test to find Accuracy
   v. Accuracy Obtained = 64.7556%

Model	**Accuracy**
SVM(Polynomial)	64.3513%
SVM(Gaussian)	64.7556%

Accuracy Table for Problem1

Problem 2

First, we load the .mat files into the workspace as X_train,y_train,X_test,y_test

a. We use a KNN classifier to classify data.into respective classes.
    i. We use the 'fitcknn' function from MATLAB to create a Model using Nearest Neighbors =7.
    ii. Then, the 'predict' function is used on X_test which gives us an output of the labels for the classes.
    iii. The label matrix is then compared to the y_test to find Accuracy

b. We use SVM classifiers to classify and predict data into respective classes.
    i. We use the 'fitcsvm' function to train different SVM for every unique class present in the dataset using kernel function as polynomial and order as 2.
    ii. Then, the 'predict' function is used on the X_test and every SVM Model and a score matrix is computed.
    iii. The SVM for a class that gives the highest prediction score, it's class is chosen as the class to associate with the input. Using this we form a label matrix. The label matrix is then compared to the y_test to find Accuracy.

c. We train an Artificial Neural Network to predict the output.
    i. We create a Neural net, with one hidden layer and 25 neurons.
    ii. We then train the network using the 'train' function, Next we check the output by passing X_test to the net and getting the index of the class using the 'vec2ind' function.
    iii. This is compared with y_test to find the accuracy.

d. For the ensemble method to find accuracy, we make a label matrix from each method where each row has a prediction for the 10 classes. We then add all three label matrix, and find the index of the maximum element in each row. Naturally the element with the maximum value will be the majority prediction. Using this we make a final label matrix and compare it with the test set.

Model	**Accuracy**
KNN	95.2322%
ANN	89.68 - 91.69%
SVM	94.4940%
Ensemble	95.1707%

Accuracy Table for Problem2